

修士論文

ミクロカノニカル純粋状態の制限ボルツマンマシンによる生成

高橋 岳志

2020 年 2 月 12 日

目次

第 1 章	序論	2
1.1	ボルツマンマシン	2
1.2	熱的量子純粋状態	5
第 2 章	目的と手法	8
2.1	本研究の目的	8
2.2	数値シミュレーションの手法	8
第 3 章	数値シミュレーションの結果	12
3.1	準備	12
3.2	結果	12
第 4 章	結論	14
	参考文献	15
	謝辞	16

第 1 章

序論

本研究で必要となった既存の概念、技術、アイデアについてまとめる。前半はボルツマンマシンについて、後半は熱的量子純粋状態について述べる。ボルツマンマシンとは機械学習の手法の一つであり、本研究ではこれを用いて数値計算を行った。熱的量子純粋状態は統計力学に現れる一つの量子論的純粋状態であるが、本研究では数値計算によりこの状態を生成した。

1.1 ボルツマンマシン

昨今、様々な領域で、機械学習の手法が成功を収めている。それは、画像認識や音声認識、あるいはイラスト生成やゲーム対戦などの多岐に渡る。この流行に乗り、物理学の分野でも、機械学習の手法で問題を解決しようということが試みられている。この節では、ボルツマンマシンについてその数理的側面を明らかにすることで、物理学への応用を考える前の準備としたい。

1.1.1 ボルツマンマシンの定義

はじめに格子系を定義しよう。 $k = 1, 2, \dots, d, i = 1, 2, \dots, N$ について $x_i^{(k)} \in \mathbb{Z}$ であるような d 次元の数ベクトル $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(d)})$ を格子点とする。ただし、 $i \neq j$ ならば $x_i \neq x_j$ を満たす。その集合を Λ とする。次に、ボンド（格子点間の結合）を Λ の元のペア $\{x, y\}$ で定義する。そして、ボンドの集合を B とする。格子系とはこの 2 つの集合の組 (Λ, B) のことである。

続いて、このような格子点上に $1, -1$ の値をとるスピン変数 σ_x を定義する。そのようなスピン変数を要素に持つ数ベクトルを σ と書くことにする。

σ を確率変数とする確率モデルを考える。 σ の取りうるすべての値の集合を Ω とかく。確率分布 $p(\sigma)$ は以下のハミルトニアンによるカノニカル分布で与えられる。

$$\mathcal{S}(\sigma) = \sum_{\{x,y\} \in B} \sigma_x w_{xy} \sigma_y + \sum_{x \in \Lambda} b_x \sigma_x \quad (1.1)$$

ここで w_{xy}, b_x はそれぞれ、重み、バイアスと呼ばれる量である。温度をこの重みとバイアスに取り入れてしまえば、カノニカル分布 $p(\sigma)$ は

$$p(\sigma) = \frac{e^{\mathcal{S}(\sigma)}}{Z} \quad (1.2)$$

と書ける。 Z は分配関数である。これは結合強度と磁場がそれぞれ重みとバイアスに一般化されたイジング模型そのものである。ボルツマンマシンとはこのようなイジング模型のことである。機械学習ではある確率分布をこのイジング模型に再現させるため、重みとバイアスを最適化する。

1.1.2 制限ボルツマンマシン

続いて、上で定義したボルツマンマシンに制限を加えて、計算上都合の良い形に書き換える。制限とは次のようなものである；格子系 (Λ, B) はバイパータイトである。バイパータイトであるとは以下のよう

に定義される。

定義 1. 格子系 (Λ, B) について、格子点 Λ を2つの集合 Λ_1, Λ_2 に分ける。そのとき、 Λ_1 の元だけ、 Λ_2 の元だけのペアには結合がなく、 Λ_1, Λ_2 の間でペアを作ったときにのみ結合がある。そのような分割 (Λ_1, Λ_2) が存在するとき、格子系 (Λ, B) はバイパータイトであるという。

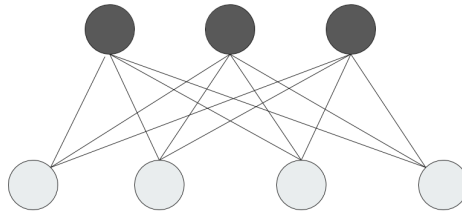


図 1.1 バイパータイトな格子系の模式図。格子点を丸で表し、ボンドを線で表している。格子点は黒丸の集合と白丸の集合に分割されていることが見て分かる。黒丸同士、白丸同士には結合がなく、黒白間にのみ結合が存在する。

制限ボルツマンマシンは、例えば、以下のようなハミルトニアンを持っている。

$$\mathcal{S}(\mathbf{v}, \mathbf{h}) = \sum_{\{x,y\} \in B} h_x w_{xy} v_y + \sum_{x \in \Lambda_2} h_x b_x + \sum_{y \in \Lambda_1} c_y v_y \quad (1.3)$$

確率変数 σ を $\mathbf{v} \in \Lambda_1, \mathbf{h} \in \Lambda_2$ に分け、それぞれの間の結合を w_{xy} で表している。機械学習の言葉ではこの Λ_1 を可視層、 Λ_2 を隠れ層と呼ぶ。可視層と隠れ層の確率変数の取りうるすべての値の集合をそれぞれ Ω_1, Ω_2 とする。

1.1.3 ブロック化ギブスサンプリング

通常、イジング模型の期待値は適当なマルコフ連鎖を生成しサンプリングをすることで計算される。そのマルコフ連鎖にエルゴード性を課せば、サンプルの時間平均と統計平均が一致する。大数の法則により、確率変数の統計平均はその期待値に収束する。マルコフ連鎖にエルゴード性をもたせるために、詳細釣り合いの条件と呼ばれるより強い性質を与えることが多い。

制限ボルツマンマシンはそのようなマルコフ連鎖を作る上で都合が良い。ある状態 \mathbf{v}, \mathbf{h} が時刻 t で実現する確率をそれぞれ $p(\mathbf{v}; t), p(\mathbf{h}; t)$ としたときの制限ボルツマンマシンのマルコフ連鎖を以下のように定義する。

定義 2. 以下のようなコルモゴロフの先進方程式に従って時間発展するマルコフ連鎖を定義する。

$$\begin{aligned} p(\mathbf{v}; t+1) &= \sum_{\mathbf{h}' \in \Omega_2} T(\mathbf{v}|\mathbf{h}')p(\mathbf{h}'; t) \\ p(\mathbf{h}; t+1) &= \sum_{\mathbf{v}' \in \Omega_1} T(\mathbf{h}|\mathbf{v}')p(\mathbf{v}'; t) \end{aligned} \quad (1.4)$$

ここで $T(\mathbf{v}|\mathbf{h}), T(\mathbf{h}|\mathbf{v})$ は遷移行列と呼ばれる条件付き確率で、

$$\begin{aligned} T(\mathbf{v}|\mathbf{h}) &= \frac{e^{\mathcal{S}(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{v}' \in \Omega_1} e^{\mathcal{S}(\mathbf{v}', \mathbf{h})}} \\ T(\mathbf{h}|\mathbf{v}) &= \frac{e^{\mathcal{S}(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{h}' \in \Omega_2} e^{\mathcal{S}(\mathbf{v}, \mathbf{h}')}} \end{aligned} \quad (1.5)$$

で表せられる。

詳細釣り合いの条件を定義しておこう。

定義 3 (詳細釣り合い). 遷移行列 T_{ij} について、等式

$$T_{ij}\Pi_j = T_{ji}\Pi_i \quad (1.6)$$

が成り立つとき、このマルコフ連鎖は詳細釣り合いの条件を満たすという。

詳細釣り合いの条件が満たされるとき、マルコフ連鎖は $t \rightarrow \infty$ で確率分布 Π_i に収束する。上で定義したマルコフ連鎖は詳細釣り合いの条件を満たす。これを示すために、補題としてベイズの定理を紹介する。

定理 4 (ベイズ). 条件付き確率 $P(A|B)$ について以下の等式が成り立つ。

$$P(A|B)p(B) = P(B|A)p(A) = p(A, B) \quad (1.7)$$

ただし、 $p(A), p(B)$ は $p(A, B)$ の周辺確率である。

定理 5. (1.4) は詳細釣り合いの条件を満たす。

証明. $p(\mathbf{v}; t)$ だけのマルコフ連鎖を考える。(1.4) を連立すれば、以下のような方程式を得る。

$$p(\mathbf{v}; t+1) = \sum_{\mathbf{v}' \in \Omega_1, \mathbf{h}' \in \Omega_2} T(\mathbf{v}|\mathbf{h}')T(\mathbf{h}'|\mathbf{v}')p(\mathbf{v}'; t) \quad (1.8)$$

このとき、遷移行列は $T(\mathbf{v}|\mathbf{v}') = \sum_{\mathbf{h}' \in \Omega_2} T(\mathbf{v}|\mathbf{h}')T(\mathbf{h}'|\mathbf{v}')$ である。次に、遷移行列をベイズの定理を用いて書き換えよう。条件付き確率の \mathbf{v} と \mathbf{h} を入れ替える。以下の確率分布

$$\Pi(\mathbf{v}) = \frac{\sum_{\mathbf{h}' \in \Omega_2} e^{\mathcal{S}(\mathbf{v}, \mathbf{h}')}}{\sum_{\mathbf{v}' \in \Omega_1, \mathbf{h}' \in \Omega_2} e^{\mathcal{S}(\mathbf{v}', \mathbf{h}')}}, \Pi(\mathbf{h}) = \frac{\sum_{\mathbf{v}' \in \Omega_1} e^{\mathcal{S}(\mathbf{v}', \mathbf{h})}}{\sum_{\mathbf{v}' \in \Omega_1, \mathbf{h}' \in \Omega_2} e^{\mathcal{S}(\mathbf{v}', \mathbf{h}')}} \quad (1.9)$$

を定義すれば、

$$\begin{aligned} T(\mathbf{v}|\mathbf{v}') &= \sum_{\mathbf{h}' \in \Omega_2} T(\mathbf{h}'|\mathbf{v}) \frac{\Pi(\mathbf{v})}{\Pi(\mathbf{h}')} T(\mathbf{v}'|\mathbf{h}') \frac{\Pi(\mathbf{h}')}{\Pi(\mathbf{v}')} \\ &= \sum_{\mathbf{h}' \in \Omega_2} T(\mathbf{v}'|\mathbf{h}') T(\mathbf{h}'|\mathbf{v}) \frac{\Pi(\mathbf{v})}{\Pi(\mathbf{v}')} \\ &= T(\mathbf{v}'|\mathbf{v}) \frac{\Pi(\mathbf{v})}{\Pi(\mathbf{v}')} \end{aligned} \quad (1.10)$$

となる。 $p(\mathbf{h}; t)$ も同様に計算できる。以上より、マルコフ連鎖 (1.4) は詳細釣り合いの条件を満たす。□

ブロック化ギブスサンプリングではマルコフ連鎖 (1.4) が生成する確率分布から \mathbf{v} と \mathbf{h} のサンプリングを交互に行い、その時間平均を計算する。しかし、ブロック化ギブスサンプリングによって得られる終分布は期待していたボルツマン因子の形をしていない。その形は $\Pi(\mathbf{v}, \mathbf{h}) = \Pi(\mathbf{v})\Pi(\mathbf{h})$ だと計算によって確かめられる。本研究ではブロック化ギブスサンプリングの代わりにメトロポリス法でのサンプリングを行った。メトロポリス法とは一様乱数で次の状態を決め、以下に定義する棄却率 α で状態を更新するかしないかを確率的に決定する。

$$\alpha(\mathbf{v}, \mathbf{h} \rightarrow \mathbf{v}', \mathbf{h}') = \min \left\{ 1, \frac{e^{\mathcal{S}(\mathbf{v}', \mathbf{h}')}}{e^{\mathcal{S}(\mathbf{v}, \mathbf{h})}} \right\} \quad (1.11)$$

このようなマルコフ連鎖は詳細釣り合いの条件をみたし、終分布は $\sim e^{\mathcal{S}(\mathbf{v}, \mathbf{h})}$ となる。

1.1.4 機械学習

「モデルのパラメータを調整することで、ある確率分布を再現すること」を、本論文では機械学習と呼ぶ。上で定義した制限ボルツマンマシンを機械学習させることを考える。

機械学習をするためには、再現したい確率分布と現在の確率分布との間の誤差を与える関数を定義する必要がある。今、再現したい確率分布が何かの実験で得られた統計データの経験分布 $q(\mathbf{v})$ だとする。このような状況で制限ボルツマンマシンを用いて機械学習を行う場合は、以下のような誤差関数を用いる。

$$E(\theta) = \sum_{\mathbf{v} \in \Omega_1} q(\mathbf{v}) \log \frac{q(\mathbf{v})}{\Pi_{\theta}(\mathbf{v})} \quad (1.12)$$

これは、誤差関数としてカルバックライブラーダイバージェンスを用いたものである。また、パラメータ w, b をまとめて θ と記述した。

上で定義した誤差関数のパラメータ空間の最小点が、我々が求めたいパラメータである。そのような最適化問題を解析的に解くのは一般には困難である。そこで機械学習では、勾配降下法を用いて誤差関数の最小点を探索する。勾配降下法とは、誤差を減らす方向に少しずつパラメータを変更することで、誤差関数の最小点を求める手法である。この手法の最も単純なアルゴリズム（最急降下法）は以下のようなものである。

$$\theta^{(n+1)} = \theta^{(n)} - \eta \nabla E(\theta) \quad (1.13)$$

現在、機械学習の分野では、上のアルゴリズムよりも高速なものが開発されている。例えば、Momentum 法、AdaGrad、RMSprop、ADAM などが存在する。

1.2 熱的量子純粋状態

熱的量子純粋状態（以下、TPQ 状態と略す）とは、ある平衡状態に対応するミクロカノニカルエネルギーシエルから一様ランダムに選んだ状態 $|\psi\rangle$ に対する力学変数 \hat{A} の量子力学的期待値 $\langle \psi | \hat{A} | \psi \rangle$ が、通常ミクロカノニカル分布で得られる力学変数の平衡値 $\langle \hat{A} \rangle_{E,N}^{MC}$ を 1 に近い確率で十分良く近似する、という性質によって定義されている [3, 4, 5]。この概念は統計力学の“典型性”によって正当化される。典型性とはマクロな量子系の平衡状態における仮定で、次のようなステートメントである；ある平衡状態に対応する状態たちのうち、大多数の状態はほとんど区別できない。要するに、膨大な数の対応する量子状態

は、マクロに見ればそっくりなもので構成されている、ということである。その“典型的な”一つ一つの状態がもつ性質が平衡状態の性質を決定する。この性質により、たった一つの状態を持ってきたとしても、その状態から熱力学的な性質を導くことが可能になる。先ほどの力学変数についての平衡値はミクロカノニカル分布での期待値と等しいことがわかっているが、力学変数以外の量—例えばエントロピーや温度—についても計算方法を与えたのが [6] である。

1.2.1 熱的量子純粋状態の定義

N サイトの量子スピン系を用意する。この系のハミルトニアンが生成するヒルベルト空間を H とし、エネルギー E をもつ固有ベクトルが張る H の部分空間を H_E とする（ミクロカノニカルエネルギーシェル）。

以下のようなステートメントで熱的量子純粋状態を定義する。

定義 6 (熱的量子純粋状態). 状態 $|\psi\rangle$ をエネルギー E のミクロカノニカルエネルギーシェルの中から一様乱数で選ぶとき、任意の正数 ϵ について、全ての力学変数 \hat{A} に対して

$$P(|\langle\psi|\hat{A}|\psi\rangle - \langle\hat{A}\rangle_{E,N}^{MC}| \geq \epsilon) \leq \eta_\epsilon(N)$$

を満たすとする。このときの $|\psi\rangle$ を熱的量子純粋状態と呼ぶ。ここで $P(x)$ は事象 x が起こる確率を表し、 $\langle\cdot\rangle_{E,N}^{MC}$ はエネルギー E 、システムサイズ N のミクロカノニカル分布によるアンサンブル平均を表す。また右辺の $\eta_\epsilon(N)$ は $N \rightarrow \infty$ で 0 になる関数である。

1.2.2 カノニカル分布の導出

全系がシステムと環境からなる格子点上で定義された量子スピン系を考える。それぞれのヒルベルト空間を H_S, H_B と書く。全系のハミルトニアンを $\hat{H} = \hat{H}_S + \hat{H}_B + \lambda\hat{H}_{\text{int}}$ で定義し、 \hat{H}_S, \hat{H}_B の固有ベクトルを $\{|\Phi_n\rangle\}_{(n=1,2,\dots)}, \{|\Psi_m\rangle\}_{(m=1,2,\dots)}$ とする。すると全ヒルベルト空間 $H = H_S \otimes H_B$ の任意のベクトルは、複素数 $c_{n,m}$ を用いて、 $|\psi\rangle = \sum_{n,m} c_{n,m} |\Phi_n\rangle |\Psi_m\rangle$ のように表される。

エネルギー E の規格化されていない TPQ 状態を以下のように定義する。

$$|E\rangle = \sum_{n,m} g_{n,m}^E |\Phi_n\rangle |\Psi_m\rangle \quad (1.14)$$

このとき $|E\rangle$ について、エネルギーの期待値と分散は以下のように計算できる。

$$\begin{aligned} U &= \frac{\langle E|\hat{H}|E\rangle}{\langle E|E\rangle} \\ &\simeq \frac{\int_{-\infty}^{\infty} u |g^E(u)|^2 D(u) du}{\int_{-\infty}^{\infty} |g^E(u)|^2 D(u) du} + O(\lambda) \\ V &\simeq \frac{\int_{-\infty}^{\infty} u^2 |g^E(u)|^2 D(u) du}{\int_{-\infty}^{\infty} |g^E(u)|^2 D(u) du} - \left\{ \frac{\int_{-\infty}^{\infty} u |g^E(u)|^2 D(u) du}{\int_{-\infty}^{\infty} |g^E(u)|^2 D(u) du} \right\}^2 + O(\lambda) \end{aligned} \quad (1.15)$$

期待値の計算をするときに、全系の状態密度 $D(u)$ を用いて和を積分に置き換えた。 $D(u)$ はボルツマンの公式により、全系のエントロピー $S(u)$ を用いて、

$$D(u) = \exp[S(u)] \quad (1.16)$$

のように表せる。そして、このエネルギーの期待値と分散について、

$$\begin{aligned} U &= E + O(\lambda) \\ V &= O(\lambda) \end{aligned} \quad (1.17)$$

が成り立つことを要請する。

次に、 $|E\rangle$ の環境成分をトレースアウトしたとき、システムはカノニカル分布になることを証明する。すなわち、以下の命題を示す。

命題 7. 熱浴とシステムの2系からなる複合系のミクロカノニカル TPQ 状態 $|E\rangle$ について、以下の関係が成り立つ。

$$\frac{\langle \Phi_n | \text{Tr}_B [|E\rangle\langle E |] | \Phi_n \rangle}{\langle E | E \rangle} \simeq \frac{e^{-\beta E_n}}{Z(\beta)} \quad (1.18)$$

証明. この分布の n 成分目を計算する。 $\langle \Phi_n | \text{Tr}_B [|E\rangle\langle E |] | \Phi_n \rangle = \sum_m |g_{nm}^E|^2$ となるが、これを積分表示すると、

$$\langle \Phi_n | \text{Tr}_B [|E\rangle\langle E |] | \Phi_n \rangle = \int_{-\infty}^{\infty} |g^E(u + E_n)|^2 D_B(u) du \quad (1.19)$$

のようになる。ここで、 $D_B(u)$ は環境系の状態密度で、 E_n はシステムの n 番目のエネルギー固有値である。また、 $D_B(u)$ は環境系のエントロピーをボルツマンの公式に用いて、 $D_B(u) = e^{S_B(u)}$ と書ける (B は Bath、つまり熱浴のことを表す)。次に被積分関数を E_n だけ右にシフトする。その後で、被積分関数に $D(u)/D(u)$ を掛け算する。ここで現れた $D_B(u - E_n)/D(u)$ という因子は、ボルツマンの公式により $e^{S_B(u - E_n) - S(u)}$ と書き表せる。これらの操作により、得られた計算結果は結局、

$$\langle \Phi_n | \text{Tr}_B [|E\rangle\langle E |] | \Phi_n \rangle = \int_{-\infty}^{\infty} |g^E(u)|^2 D(u) e^{S_B(u - E_n) - S(u)} du \quad (1.20)$$

というものになる。ところで、この節での要請により、 $|g^E(u)|^2 D(u)$ は $u = E$ で鋭いピークを持つ関数である。続いて、この積分を分母を含めて考え、ゾンマーフェルト展開する。結果は以下ようになる。

$$\begin{aligned} \frac{\langle \Phi_n | \text{Tr}_B [|E\rangle\langle E |] | \Phi_n \rangle}{\langle E | E \rangle} &= \frac{e^{S_B(E - E_n) - S(E)}}{\int_{-\infty}^{\infty} |g^E(u)|^2 D(u) du} + I_1 + I_2 + O(\lambda) \\ I_1 &= \left(\frac{d}{dE} e^{S_B(E - E_n) - S(E)} \right) \frac{\int_{-\infty}^{\infty} (u - E) |g^E(u)|^2 D(u) du}{\int_{-\infty}^{\infty} |g^E(u)|^2 D(u) du} \\ I_2 &= \left(\frac{1}{2!} \frac{d^2}{dE^2} e^{S_B(E - E_n) - S(E)} \right) \frac{\int_{-\infty}^{\infty} (u - E)^2 |g^E(u)|^2 D(u) du}{\int_{-\infty}^{\infty} |g^E(u)|^2 D(u) du} \end{aligned} \quad (1.21)$$

I_1, I_2 に出てくる積分は、この節での要請から $\simeq O(\lambda)$ となることがわかる。これより、 λ の一次のオーダーで、

$$\frac{\langle \Phi_n | \text{Tr}_B [|E\rangle\langle E |] | \Phi_n \rangle}{\langle E | E \rangle} \simeq \frac{e^{S_B(E - E_n) - S(E)}}{\int_{-\infty}^{\infty} |g^E(u)|^2 D(u) du} \quad (1.22)$$

が成り立つ。ここで熱力学を用いる。全系のエントロピー $S(E)$ は平衡状態の部分系のエントロピーを用いて、 $S(E) = S_S(E) + S_B(E)$ と分解できる。 $S_S(E)$ の方は E_n に依存しないので分母に置いておく。そこで、 $S_B(E - E_n) - S_B(E)$ を計算する。エントロピーをエネルギーと自由エネルギーについてのルジャンドル変換で表すと以下ようになる。

$$S(U) = \inf_{\beta} \{ \beta U - F(\beta) \} \quad (1.23)$$

この式を用いれば、 $S_B(E - E_n) - S_B(E) = -\beta E_n$ であることがわかる。ここで β は上のルジャンドル変換、あるいは $\beta(u) = \partial S(u) / \partial u$ によって決まる逆温度である。以上より命題が証明された。 \square

第 2 章

目的と手法

前節で紹介したボルツマンマシンが、熱的量子純粋状態を生成するように機械学習を行う。はじめに、熱的量子純粋状態を学習したボルツマンマシンからどのような情報を引き出したいか、ということを説明する。その次に、本研究で実行した数値シミュレーションのプログラムについて主に解説する。コーディングには Julia 言語を用いた。

2.1 本研究の目的

熱的量子純粋状態を学習した制限ボルツマンマシンは、(1.14) の波動関数に対応する確率分布 $|g_{n,m}^E|^2$ で状態を発生させる古典的確率モデルになっている。しかし、一般に波動関数は複素数である。複素数値を学習させるためには、重みとバイアスを複素数に拡張した、複素ボルツマンマシンを使わなくてはならない。複素ボルツマンマシンでは遷移行列として、 $\propto |e^{S(\mathbf{v}, \mathbf{h})}|^2$ となるようなものを採用し、マスター方程式を解く。この際、 $e^{S(\mathbf{v}, \mathbf{h})}$ が熱的量子純粋状態の波動関数に比例するように最適化を行う。

この複素ボルツマンマシンによって得られる熱的量子純粋状態について部分トレースをとることでシステムについてのカノニカル分布を得る。制限ボルツマンマシンはこのような部分トレースが容易に計算できる。このようにして得られたカノニカル分布から力学変数を計算する。本研究では 2 スピンが相互作用する量子スピン系についてエネルギーを計算し、理論値と比較した。

2.2 数値シミュレーションの手法

2.2.1 誤差関数について

通常の機械学習では実験や調査などで得られた統計データを再現する確率分布を学習するが、本研究ではそのような統計データは存在しない。その代わり、TPQ 状態はエネルギー E にシャープなピークを持つ状態として定義されている (1.17) ので、以下のような誤差関数を最小化すればよい。

$$\text{Error}(E; \theta) = (U(\theta) - E)^2 \quad (2.1)$$

$U(\theta)$ は以下のような計算によって求められるエネルギー期待値である。

$$U(\theta) = \sum_{\mathbf{v}, \mathbf{v}' \in \Omega_1, \mathbf{h}, \mathbf{h}' \in \Omega_2} e^{S^*(\mathbf{v}, \mathbf{h})} \mathcal{H}(\mathbf{v}, \mathbf{h}, \mathbf{v}', \mathbf{h}') e^{S(\mathbf{v}', \mathbf{h}')} / Z \quad (2.2)$$

ただし、 $\mathcal{H}(\mathbf{v}, \mathbf{h}, \mathbf{v}', \mathbf{h}')$ は計算したい系のハミルトニアン行列の行列要素である。このような複素数値の

サンプリング、そして行列のサンプリングには、通常とは異なる手法が必要である。次のセクションでは期待値の計算手法について述べる。

2.2.2 変分モンテカルロ法

遷移行列を以下のように定義する。

$$\begin{aligned} T(\mathbf{v}|\mathbf{h}) &= \frac{|e^{\mathcal{S}(\mathbf{v},\mathbf{h})}|^2}{\sum_{\mathbf{v}' \in \Omega_1} |e^{\mathcal{S}(\mathbf{v}',\mathbf{h})}|^2} \\ T(\mathbf{h}|\mathbf{v}) &= \frac{|e^{\mathcal{S}(\mathbf{v},\mathbf{h})}|^2}{\sum_{\mathbf{h}' \in \Omega_2} |e^{\mathcal{S}(\mathbf{v},\mathbf{h}')}|^2} \end{aligned} \quad (2.3)$$

この遷移行列に従って時間発展したマスター方程式から確率変数をサンプリングする。しかし、マスター方程式の収束先は波動関数を直接表現しているわけではないので工夫が必要である。ここでは変分モンテカルロ法 [7, 8] と呼ばれる手法を用いる。

変分モンテカルロ法とは量子系のような行列要素をサンプリングするときに用いられるモンテカルロ法の一つである。量子系のエネルギー期待値の式を変形することで、行列要素のサンプリングを通常のモンテカルロ法によるものに帰着させる。エネルギー期待値 $\sum_{\mathbf{v},\mathbf{h},\mathbf{v}',\mathbf{h}'} \psi^*(\mathbf{v},\mathbf{h}) \mathcal{H}(\mathbf{v},\mathbf{h},\mathbf{v}',\mathbf{h}') \psi(\mathbf{v}',\mathbf{h}')$ を以下のような形に変形する。

$$\sum_{\mathbf{v},\mathbf{h},\mathbf{v}',\mathbf{h}'} \psi^*(\mathbf{v},\mathbf{h}) \mathcal{H}(\mathbf{v},\mathbf{h},\mathbf{v}',\mathbf{h}') \psi(\mathbf{v}',\mathbf{h}') = \sum_{\mathbf{v},\mathbf{h}} \left\{ \sum_{\mathbf{v}',\mathbf{h}'} \mathcal{H}(\mathbf{v},\mathbf{h},\mathbf{v}',\mathbf{h}') \frac{\psi(\mathbf{v}',\mathbf{h}')}{\psi(\mathbf{v},\mathbf{h})} \right\} |\psi(\mathbf{v},\mathbf{h})|^2 \quad (2.4)$$

そこで、 $\sum_{\mathbf{v}',\mathbf{h}'} \mathcal{H}(\mathbf{v},\mathbf{h},\mathbf{v}',\mathbf{h}') \psi(\mathbf{v}',\mathbf{h}') / \psi(\mathbf{v},\mathbf{h})$ を確率変数だと思って $|\psi(\mathbf{v},\mathbf{h})|^2$ からサンプリングを行う。ここで現れた確率変数をローカルエネルギーと呼ぶ。

システムだけの確率変数の期待値をマルコフ連鎖から計算するとき、制限ボルツマンマシンでは $\Pi(\mathbf{v},\mathbf{h})$ の周辺分布からのサンプリングを行っていることになる。制限ボルツマンマシンは古典系なので、周辺化と部分トレースが一致する。序論で述べたとおり、熱的量子純粋状態の部分トレースはカノニカル分布になるので、結果的にカノニカル分布からのサンプリングになる。

サンプリングの手法を記しておく。 v_x, h_y がそれぞれ ± 1 をとる場合を考える。条件付き確率では、 \mathbf{h} の値が決められた状態で \mathbf{v} のサイコロをふる。ボルツマンマシンのハミルトニアン (1.1) を見るとわかるように、ある \mathbf{h} が決まった場合、異なる v_x は独立に計算できる。これをボルツマンマシンの条件付き独立性と呼ぶ。

ソースコード 2.1 ブロック化ギブスサンプリング

```

1 s = -ones(Float64, Const.dimS)
2 prob = 1.0 ./ (1.0 .+ exp.(-2.0 * z))
3 pup = rand(Float64, Const.dimS)
4 for ix in 1:Const.dimS
5     if pup[ix] < prob[ix]
6         s[ix] = 1.0
7     end
8 end
9 return s

```

ここで s はサイコロの出目、 Const.dimS はスピンのサイト数である。 z は $\sum_y w_{xy} h_y$ をまとめて書いたもので、これは機械学習の言葉では「活性」と呼ばれる。条件付き独立性のため、異なる格子点に対応する乱数を乱数発生関数を用いて一度に計算することができる。

2.2.3 最適化手法

最適化を行うために、上で計算された期待値の微分をとらなくてはならない。(2.2) について、 $\theta = (w, b, c)$ の微分は以下ようになる。

$$\begin{aligned}\partial_{\theta^*} U(\theta) &= \sum_{\mathbf{v}, \mathbf{v}' \in \Omega_1, \mathbf{h}, \mathbf{h}' \in \Omega_2} e^{\mathcal{S}^*(\mathbf{v}, \mathbf{h})} \partial_{\theta^*} \mathcal{S}^*(\mathbf{v}', \mathbf{h}') \mathcal{H}(\mathbf{v}, \mathbf{h}, \mathbf{v}', \mathbf{h}') e^{\mathcal{S}(\mathbf{v}', \mathbf{h}')}/Z - \\ &\quad \sum_{\mathbf{v}, \mathbf{v}' \in \Omega_1, \mathbf{h}, \mathbf{h}' \in \Omega_2} e^{\mathcal{S}^*(\mathbf{v}, \mathbf{h})} \partial_{\theta^*} \mathcal{S}^*(\mathbf{v}', \mathbf{h}') e^{\mathcal{S}(\mathbf{v}', \mathbf{h}')}/Z \times \\ &\quad \sum_{\mathbf{v}, \mathbf{v}' \in \Omega_1, \mathbf{h}, \mathbf{h}' \in \Omega_2} e^{\mathcal{S}^*(\mathbf{v}, \mathbf{h})} \mathcal{H}(\mathbf{v}, \mathbf{h}, \mathbf{v}', \mathbf{h}') e^{\mathcal{S}(\mathbf{v}', \mathbf{h}')}/Z \\ &= \langle \hat{\mathcal{H}} \partial_{\theta^*} \hat{\mathcal{S}}^* \rangle - \langle \hat{\mathcal{H}} \rangle \langle \partial_{\theta^*} \hat{\mathcal{S}}^* \rangle\end{aligned}\tag{2.5}$$

よって誤差関数の微分は

$$\partial_{\theta^*} \text{Error}(E; \theta) = 2(\langle \hat{\mathcal{H}} \rangle - E)(\langle \hat{\mathcal{H}} \partial_{\theta^*} \hat{\mathcal{S}}^* \rangle - \langle \hat{\mathcal{H}} \rangle \langle \partial_{\theta^*} \hat{\mathcal{S}}^* \rangle)\tag{2.6}$$

である。

このようにして計算された誤差関数の微分を用いて勾配降下法を行う。本研究で用いた手法は通常の最急降下法より高速な Momentum である。Momentum では勾配の他に、モーメントと呼ばれる量を計算しパラメータの最適化を図る。以下に例として重みを Momentum によって更新するコードをのせる。

ソースコード 2.2 Momentum

```
1 wmoment = 0.9 * wmoment - lr * dweight
2 weight += wmoment
```

$dweight, wmoment$ はそれぞれ勾配、モーメントである。

2.2.4 温度の計算

カノニカル分布を求めるためには温度の計算をしなくてはならない。今、熱浴がシステムと比べて非常に大きいときを考える。すると熱浴はミクロカノニカル分布になっていると考えられる。ミクロカノニカル分布でのエントロピーを用いて、温度を計算することができる。

熱浴として、ハードコアボソンの系を考える。以下のようなハミルトニアンを考える。ただし $N+1=1$ となるような周期境界条件を課す。

$$\hat{\mathcal{H}}_B = -t \sum_{y=1}^N \hat{b}_y^\dagger \hat{b}_{y+1} + \hat{b}_y \hat{b}_{y+1}^\dagger - \mu \sum_{y=1}^N \hat{b}_y^\dagger \hat{b}_y\tag{2.7}$$

$\hat{b}_y, \hat{b}_y^\dagger$ はハードコアボソンの生成消滅演算子で、以下の交換関係と反交換関係を満たす。

$$\begin{aligned}[\hat{b}_y, \hat{b}_{y'}^\dagger] &= [\hat{b}_y, \hat{b}_{y'}] = [\hat{b}_y^\dagger, \hat{b}_{y'}^\dagger] = 0 \quad (y \neq y') \\ \{\hat{b}_y, \hat{b}_y^\dagger\} &= \{\hat{b}_y, \hat{b}_y\} = 0 \quad \{\hat{b}_y, \hat{b}_y^\dagger\} = 1\end{aligned}\tag{2.8}$$

このハミルトニアンはスピン 1/2 の XY 模型

$$\hat{\mathcal{H}}_{XY} = 2t \sum_{y=1}^N \hat{S}_y^{(1)} \hat{S}_{y+1}^{(1)} + \hat{S}_y^{(2)} \hat{S}_{y+1}^{(2)} - \mu \sum_{y=1}^N \hat{S}_y^{(3)} + \frac{1}{2} \quad (2.9)$$

と等価である。ここで $\hat{S}_y^{(i)}$ はスピン 1/2 の角運動量演算子である。この XY 模型は厳密に解け、以下のような自由エネルギーを持つ。

$$F(T) = -T \sum_k \log(\cosh(|t \cos(k) - \mu|/T)) \quad (2.10)$$

この自由エネルギーをルジャンドル変換することで、エントロピーが求まる。

$$S(U) = \min_T \frac{U - F(T)}{T} \quad (2.11)$$

またこのとき右辺を最小化させるような温度 T が求めたい温度である。

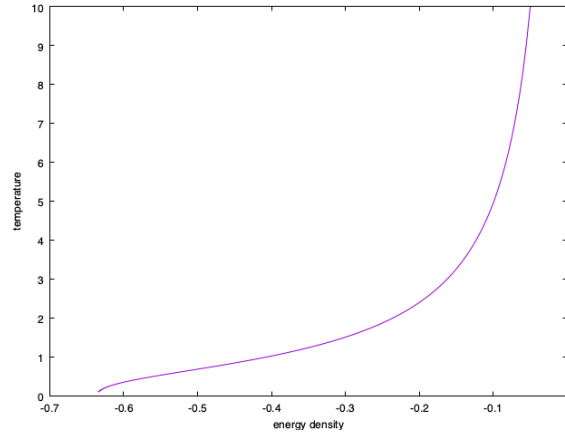


図 2.1 ハードコアボソンに対するエネルギーと温度の関係。系が基底エネルギーをとるとき温度は 0 になる。また、全ての状態が一様ランダムに出現するとき温度は ∞ に発散する。

2.2.5 初期化

パラメータ w_{xy}, b_x, c_y を初期時刻で温度が無限大に近くなるように初期化する。まず、バイアスは全ての要素が 0 のベクトルにとる。そして重みの実数成分は高温を反映して小さい量にとる。重みの虚数成分は $\sum_{h \in \Omega_2} e^{h(Wv + W^*v')} \sim \delta_{v,v'}$ になるようにとる。実際には以下のようなコードで重みの初期化を行った。

ソースコード 2.3 Initialization

```
1 weight = ones(Complex{Float64}, Const.dimB, Const.dimS) * 10.0^(-3)
2 for n in 1:Int64(Const.dimB/Const.dimS)-1
3     weight[Const.dimS*n+1:(n+1)*Const.dimS, :] +=
4     diagm(0 => ones(Complex{Float64}, Const.dimS)) * (im * pi/2.0)
5 end
```

第 3 章

数値シミュレーションの結果

3.1 準備

はじめに系の設定とシミュレーションのパラメータを記しておく。まず、具体的な系のハミルトニアンを書き、系のセットアップを定める。次にボルツマンマシンのパラメータ（機械学習ではハイパーパラメータと呼ばれる）を載せる。

3.1.1 系の設定

熱浴と部分系のハミルトニアンとして以下のものを用意した。熱浴の方には周期境界条件を課した。

$$\begin{aligned}\hat{\mathcal{H}}_B &= -t \sum_{y=1}^N \hat{b}_y^\dagger \hat{b}_{y+1} + \hat{b}_y \hat{b}_{y+1}^\dagger - \mu \sum_{y=1}^N \hat{b}_y^\dagger \hat{b}_y \\ \hat{\mathcal{H}}_S &= -J \sum_{x=1,3,5,\dots}^{M-1} \hat{S}_x \cdot \hat{S}_{x+1}\end{aligned}\tag{3.1}$$

ここで $\hat{b}_y, \hat{b}_y^\dagger$ は (2.8) の交換関係・反交換関係を満たすハードコアボソンの生成消滅演算子で、 $\hat{S}_x^{(i)}$ はスピン $1/2$ の角運動量演算子である。相互作用の定数は $t = 1, J = 1$ と設定した。

3.1.2 ハイパーパラメータ

まずモンテカルロ法に必要なパラメータを用意する。マルコフ連鎖の初期に生じる確率分布の偏りを無視するために、バーンインタイムを設けた。その回数は 100 回に設定した。そしてバーンインタイムの後、実際にサンプリングをする時間を 500 回とした。

次に学習に関する定数を設定する。最適化に至るまでの一連のマルコフ連鎖を繰り返す回数を 3000 回と定めた。また、勾配降下法でパラメータをどの程度変化させるかという意味をもつ学習率を、 1.0×10^{-7} とした。

3.2 結果

逆温度に対するシステムのエネルギー期待値の変化の様子をサイト数を変えてグラフにまとめた。

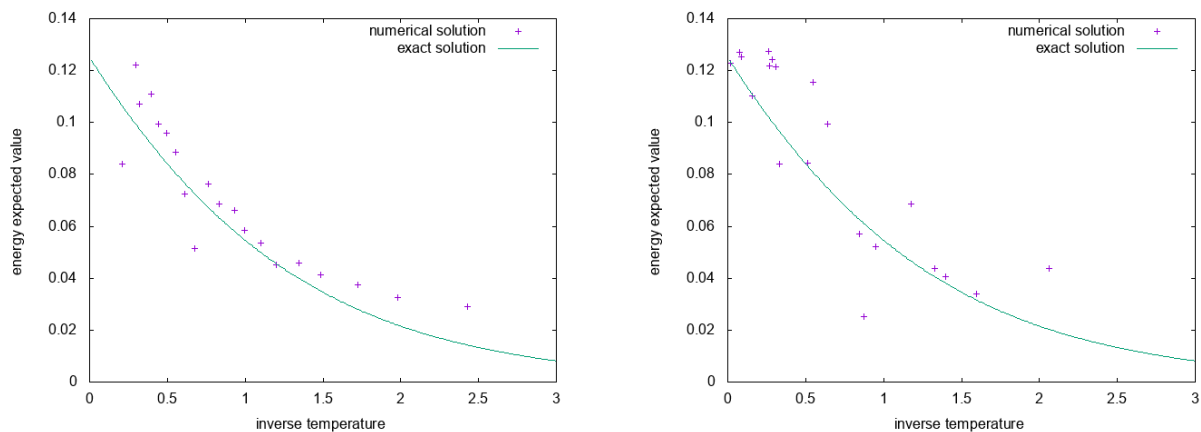


図 3.1 左図はシステムサイズ 8、熱浴サイズ 80。右図はシステムサイズ 8、熱浴サイズ 40。実線は厳密解。左図では熱浴がうまく機能していないことがわかる。

熱浴のサイズが小さいとばらつきが大きくなるということがわかる。これは、熱浴としてのサイズが小さすぎるため、システムが相対的に大きな影響を及ぼしたことが考えられる。熱浴を大きくすると、ばらつきが小さくなっていることがわかる。

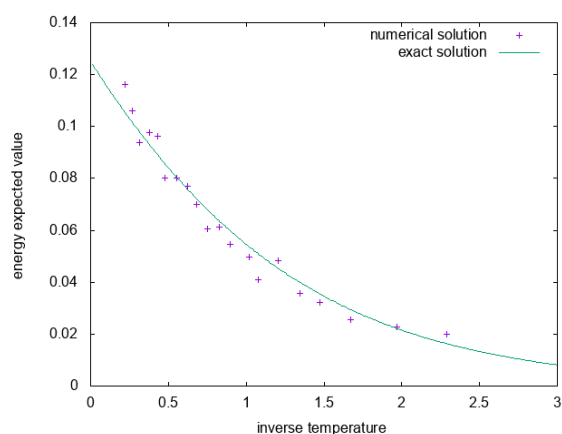


図 3.2 システムサイズ 8、熱浴サイズ 120。厳密解の曲線を中心にデータが分布している。

第 4 章

結論

シミュレーションの結果から、制限ボルツマンマシンは今回使ったモデル（ハードコアボソンとハイゼンベルク 2 スピンの複合系）の熱的純粋状態を生成できたということが結論できる。また部分系を見れば、カノニカル分布が得られるということもわかる。この結果が他のどのような系で適用できるか、拡張を試みる必要がある。

最適化には機械学習の手法を用いたが、今回の手法では虚数方向の最適化ができないことがわかった。今回は初期条件の虚数成分のパラメータを手で調整し最適化を行った。しかし、他の系ではこの操作をすることが難しい場合もありうる。そのため虚数方向の最適化を実装することをまず先に着手しなくてはならないだろう。

より高度なモデルでは制限ボルツマンマシンの表現能力を超える場合が考えられる。このようなときでも波動関数を生成できるようにするために、人工ニューラルネットワークのモデルや深層ボルツマンマシンを使うことが必要になってくるかもしれない。この場合でも周辺確率が計算できるようにするために、工夫しなくてはならないだろう。

参考文献

- [1] G. Carleo and M. Troyer, *Science* **355**, 602(2017).
- [2] H. Saito and M. Kato, *J. Phys. Soc. Jpn.* **87**, 014001(2018).
- [3] P. Bocchieri and A. Loinger, *Phys. Rev.* **114**, 948 (1959).
- [4] A. Hams and H. D. Raedt, *Phys. Rev. E* **62**, 4365 (2000).
- [5] S. Popescu, A. J. Short, and A. Winter, *Nat. Phys.* **2**, 754 (2006).
- [6] S. Sugiura and A. Shimizu, *Phys. Rev. Lett.* **108**, 240401 (2012).
- [7] D. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **16**, 3081 (1977).
- [8] H. Yokoyama and H. Shiba: *J. Phys. Soc. Jpn.* **59**, 3669 (1990).
- [9] 瀧 雅人 (2017)『機械学習スタートアップシリーズ これならわかる深層学習入門』 講談社
- [10] 田中 章詞・富谷 昭夫・橋本 幸士 (2019)『ディープラーニングと物理学 原理がわかる、応用ができる』 講談社
- [11] 橋本 幸士 編 (2019)『物理学者，機械学習を使う ―機械学習・深層学習の物理学への応用―』 朝倉書店
- [12] 田崎 晴明 (2000)『熱力学―現代的な視点から (新物理学シリーズ)』 培風館
- [13] 清水 明 (2007)『熱力学の基礎』 東京大学出版会
- [14] 田崎 晴明 (2008)『統計力学 〈1〉 (新物理学シリーズ)』 培風館
- [15] 田崎 晴明 (2008)『統計力学 〈2〉 (新物理学シリーズ)』 培風館
- [16] 小口武彦 (1970)『磁性体の統計理論』 裳華房
- [17] 田崎 晴明 『数学 物理を学び楽しむために』

謝辞

この論文を手掛ける上で多くのアドバイスを頂いた指導教官の森弘之先生には大変感謝している。また、この研究について有意義な議論をしてくださった電気通信大学情報理工学研究科基盤理工学専攻教授斎藤弘樹先生にはこの場で感謝の意を表する。最後に、一緒に議論をしてもらった研究室のメンバー、生活面で私を支えてくれた家族、そして精神的な面で私を支えてくれた鈴木美葉氏には心から感謝している。